Preliminary Results on Masquerader Detection using Compression Based Similarity Metrics

Maximiliano Bertacchini [1] Pablo I. Fierens [2] (PhD)

[1]CITEFA - Instituto de Investigaciones Científicas y Técnicas de las FFAA

[2]ITBA - Instituto Tecnológico de Buenos Aires

ASAI2006 - VIII Argentine Symposium on Artificial Intelligence
35 JAIIO - Jornadas de Informática e Investigación Operativa
Mendoza, Argentina
4 y 5 de Septiembre de 2006

Outline

Introducción

Marco

Proyecto de investigación del Laboratorio SI6 de CITEFA

- Identificación y clasificación de intrusos informáticos basadas en su comportamiento (PICTO 2004)

Objetivo de este trabajo

- Detección de intrusos en el *historial de comandos* `UNIX`
- Aplicación de la *técnica* desarrollada por M. Li et al. sobre los *datos* recolectados por M. Schonlau et al.
- Comparación de los resultados

Detección de intrusos

Clasificación de técnicas

- Misuse detection
- Anomaly detection

Detección de intrusos

Clasificación de técnicas

- Misuse detection
- Anomaly detection

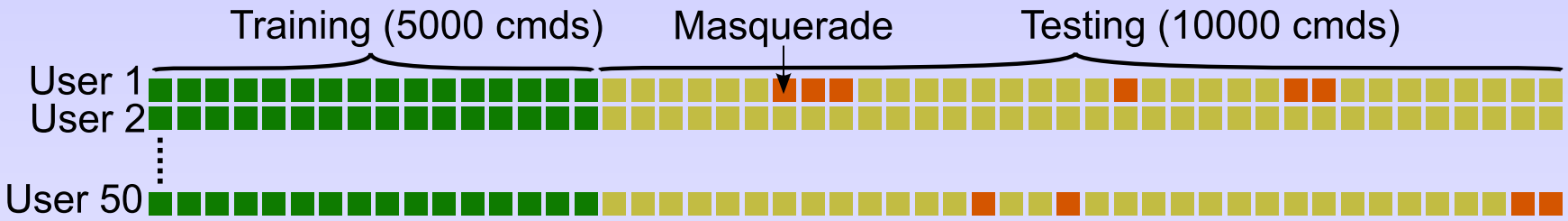
Definición

Un masquerader o *impostor* es un intruso que intenta tomar la identidad de un usuario legítimo de un sistema informático.

Trabajo previo

Schonlau et al, *"Computer Intrusion: Detecting Masquerades"*,2001

Datos *[Schonlau et al.]*[1]



Observaciones

- Fuente: `acct` (`UNIX`) (*comandos sin argumentos*)
- 50 usuarios "legítimos" + 20 "masqueraders"
- Bloques de 100 comandos
- *Testing*: un bloque es masquerader con probabilidad 1 %, y en este caso el siguiente lo es con 80 %
- Aprox. 5 % de los bloques son masqueraders
- No todos los usuarios contienen masqueraders

User 1

```
cpp
sh
xrdb
cpp
sh
xrdb
mkpts
env
csh
csh
csh
sh
kill
userenv
wait4wm
...
```

[1] `http://www.schonlau.net`

Experimentos

Metodología *[Schonlau et al.]*

- Se calcula un puntaje para cada bloque de *testing*.
- Si el puntaje supera cierto umbral, se considera **masquerader**.

Técnicas *[Schonlau et al.]*

- *Uniqueness*
- *Bayes one-step Markov*
- *Hybrid multistep Markov*
- *Sequence-Match*
- *IPAM*
- *Compression* (basado en `compress` de `UNIX`)

Compression Method *[Schonlau et al.]*

Puntaje *[Schonlau et al.]*

El **puntaje** x de Compression se define como el número adicional de bytes necesarios para comprimir un bloque de *testing* concatenado con los datos de *training*:

$$x = compress(D_{Tr}D_{Te}) - compress(D_{Tr})$$

donde D_{Tr} es training data, D_{Te} testing data, $D_{Tr}D_{Te}$ es la concatenación de ambos y $compress()$ es una función que devuelve el número de bytes de los datos comprimidos.

Compression Method *[Schonlau et al.]*

Puntaje *[Schonlau et al.]*

El **puntaje** x de Compression se define como el número adicional de bytes necesarios para comprimir un bloque de *testing* concatenado con los datos de *training*:

$$x = compress(D_{Tr}D_{Te}) - compress(D_{Tr})$$

donde D_{Tr} es training data, D_{Te} testing data, $D_{Tr}D_{Te}$ es la concatenación de ambos y $compress()$ es una función que devuelve el número de bytes de los datos comprimidos.

Umbral *[Schonlau et al.]*

Se calcula un **umbral** por usuario basado en los datos del usuario más un desplazamiento basado en los datos de todos los usuarios.

Normalized Compression Distance

Diseño

- Se reprodujo el experimento basado en `compress`, utilizando *NCD* en su lugar.
- Se intentó reproducir lo más exactamente posible el diseño de los experimentos de Schonlau et al.

Concepto

- Medida de distancia normalizada, genérica y cuasi-universal
- Basada en noción de *Kolmogorov Complexity* (no computable)
- Aproximación basada en compresores estándares

Trabajos previos

- M. Li et al, *"The Similarity Metric"*, 2004
- R. Cilibrasi et al, *"Clustering by Compression"*, 2005

Outline

Kolmogorov Complexity

Introducción...

Definición informal

Es la medida *absoluta* de los recursos computacionales *mínimos* necesarios para describir a un objeto (ej. una cadena de bits).

Ejemplo

0101010101010101010101010101010101010101010101010101...

Descripción:

Kolmogorov Complexity

Introducción...

Definición informal

Es la medida *absoluta* de los recursos computacionales *mínimos* necesarios para describir a un objeto (ej. una cadena de bits).

Ejemplo

0101010101010101010101010101010101010101010101010101...

Descripción: *"50 veces* 01*"*

Kolmogorov Complexity

Introducción...

Definición informal

Es la medida *absoluta* de los recursos computacionales *mínimos* necesarios para describir a un objeto (ej. una cadena de bits).

Ejemplo

0101010101010101010101010101010101010101010101010101...

Descripción: *"50 veces* 01*"*

Ejemplo

1100100011100001110111101110110011111010010000100101111...

Descripción:

Kolmogorov Complexity

Introducción...

Definición informal

Es la medida *absoluta* de los recursos computacionales *mínimos* necesarios para describir a un objeto (ej. una cadena de bits).

Ejemplo

0101010101010101010101010101010101010101010101010101...

Descripción: *"50 veces* 01*"*

Ejemplo

1100100011100001110111101110110011111010010000100101111...

Descripción: *¿?*

Kolmogorov Complexity

Más formalmente...

Definición

La complejidad de Kolmogorov de un objeto es la longitud de su descripción más corta en un lenguaje predefinido:

$$K(x) = min|P| : U(P) = x$$

donde P es un programa que "genera" x sin parámetros y termina, y U es una Máquina Universal de Turing.

Kolmogorov Complexity

Más formalmente...

Definición

La complejidad de Kolmogorov de un objeto es la longitud de su descripción más corta en un lenguaje predefinido:

$$K(x) = min|P| : U(P) = x$$

donde P es un programa que "genera" x sin parámetros y termina, y U es una Máquina Universal de Turing.

Teorema

$K()$ es no computable.

ITBA
INSTITUTO TECNOLOGICO DE BUENOS AIRES
UNIVERSIDAD PRIVADA

Kolmogorov Complexity

Más formalmente...

Definición

La complejidad de Kolmogorov de un objeto es la longitud de su descripción más corta en un lenguaje predefinido:

$$K(x) = min|P| : U(P) = x$$

donde P es un programa que "genera" x sin parámetros y termina, y U es una Máquina Universal de Turing.

Teorema

$K()$ es no computable.

$K(x)$ se puede aproximar "desde arriba" mediante un compresor C:

$$|K(x)| < |C(x)| < |x|$$

Compresores

Definición *[Li et al.]*

Un compresor es una función $c : \Omega \rightarrow (0,1)^*$

- Por conveniencia, definimos C como: $C(x) = |c(x)|$.
- Sólo se consideran compresores tales que $C(x) \leqslant |x| + O(\log |x|)$

Compresores

Definición *[Li et al.]*

Un compresor es una función $c : \Omega \to (0,1)^*$

- Por conveniencia, definimos C como: $C(x) = |c(x)|$.
- Sólo se consideran compresores tales que $C(x) \leqslant |x| + O(\log |x|)$

Definición *[Li et al.]*

C es un compresor normal si satisface, hasta un término aditivo $O(\log |x|)$:

- $C(xx) = C(x)$ y $C(\lambda) = 0$ donde λ es la cadena vacía (*idempotencia*)
- $C(x) \leqslant C(xy)$ (*monotonicidad*)
- $C(xy) = C(yx)$ (*simetría*)
- $C(xy) + C(z) \leqslant C(xz) + C(yz)$ (*distributividad*)

Normalized Compression Distance

Definición *[Li et al.]*

La Normalized Compression Distance entre dos cadenas x e y se define como:

$$NCD(x,y) = \frac{C(xy) - min(C(x), C(y))}{max(C(x), C(y))} \leqslant 1 + \epsilon$$

Normalized Compression Distance

Definición *[Li et al.]*

La Normalized Compression Distance entre dos cadenas x e y se define como:

$$NCD(x,y) = \frac{C(xy) - min(C(x), C(y))}{max(C(x), C(y))} \leqslant 1 + \epsilon$$

Observación

- $NCD(x,y) \to 0$ *cuando la similitud entre* x *e* y *es máxima.*
- $NCD(x,y) \to 1$ *cuando la similitud entre* x *e* y *es mínima.*

Normalized Compression Distance

Definición *[Li et al.]*

La Normalized Compression Distance entre dos cadenas x e y se define como:

$$NCD(x,y) = \frac{C(xy) - min(C(x), C(y))}{max(C(x), C(y))} \leqslant 1 + \epsilon$$

Observación

- $NCD(x,y) \to 0$ *cuando la similitud entre* x *e* y *es máxima.*
- $NCD(x,y) \to 1$ *cuando la similitud entre* x *e* y *es mínima.*

Teorema *[Cilibrasi et al, 2005]*

Si C es *normal*, entonces $NCD(x,y)$ es una similarity metric, y es cuasi-universal.

Outline

Implementación

Puntaje *NCD*

El **puntaje** de un bloque de testing se calcula como:

$$x_{NCD} = \frac{C(D_{Tr}D_{Te}) - C(D_{Te})}{C(D_{Tr})} \leqslant 1 + \epsilon$$

donde D_{Tr} es training data, D_{Te} testing data, $D_{Tr}D_{Te}$ es la concatenación de ambos y $C()$ es un compresor normal.

Implementación

Puntaje *NCD*

El **puntaje** de un bloque de testing se calcula como:

$$x_{NCD} = \frac{C(D_{Tr}D_{Te}) - C(D_{Te})}{C(D_{Tr})} \leqslant 1 + \epsilon$$

donde D_{Tr} es training data, D_{Te} testing data, $D_{Tr}D_{Te}$ es la concatenación de ambos y $C()$ es un compresor normal.

Umbral *NCD*

Se calcula un **umbral** por usuario basado en los datos del usuario más un desplazamiento basado en los datos de todos los usuarios.

ITBA
INSTITUTO TECNOLOGICO DE BUENOS AIRES
UNIVERSIDAD PRIVADA

Herramientas

CompLearn

- CompLearn[a] es una implementación de *NCD* de sus mismos autores.

[a] `http://www.complearn.org`

Compresores implementados *(normales)*

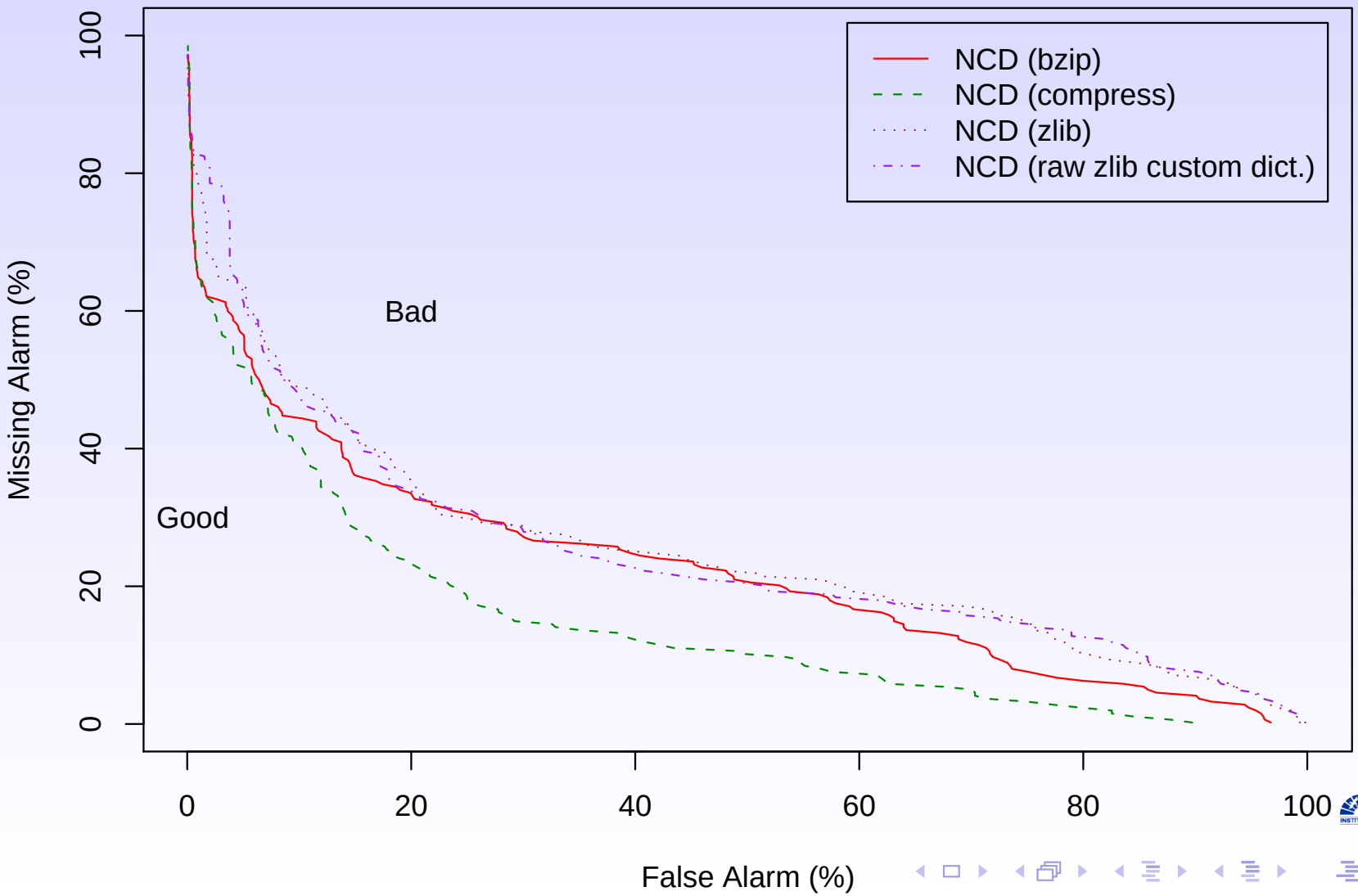
- `compress`
- `bzip2`
- `zlib`
- `raw zlib con diccionario` (compuesto por los comandos más comunes del juego de datos)

Resultados: comparación

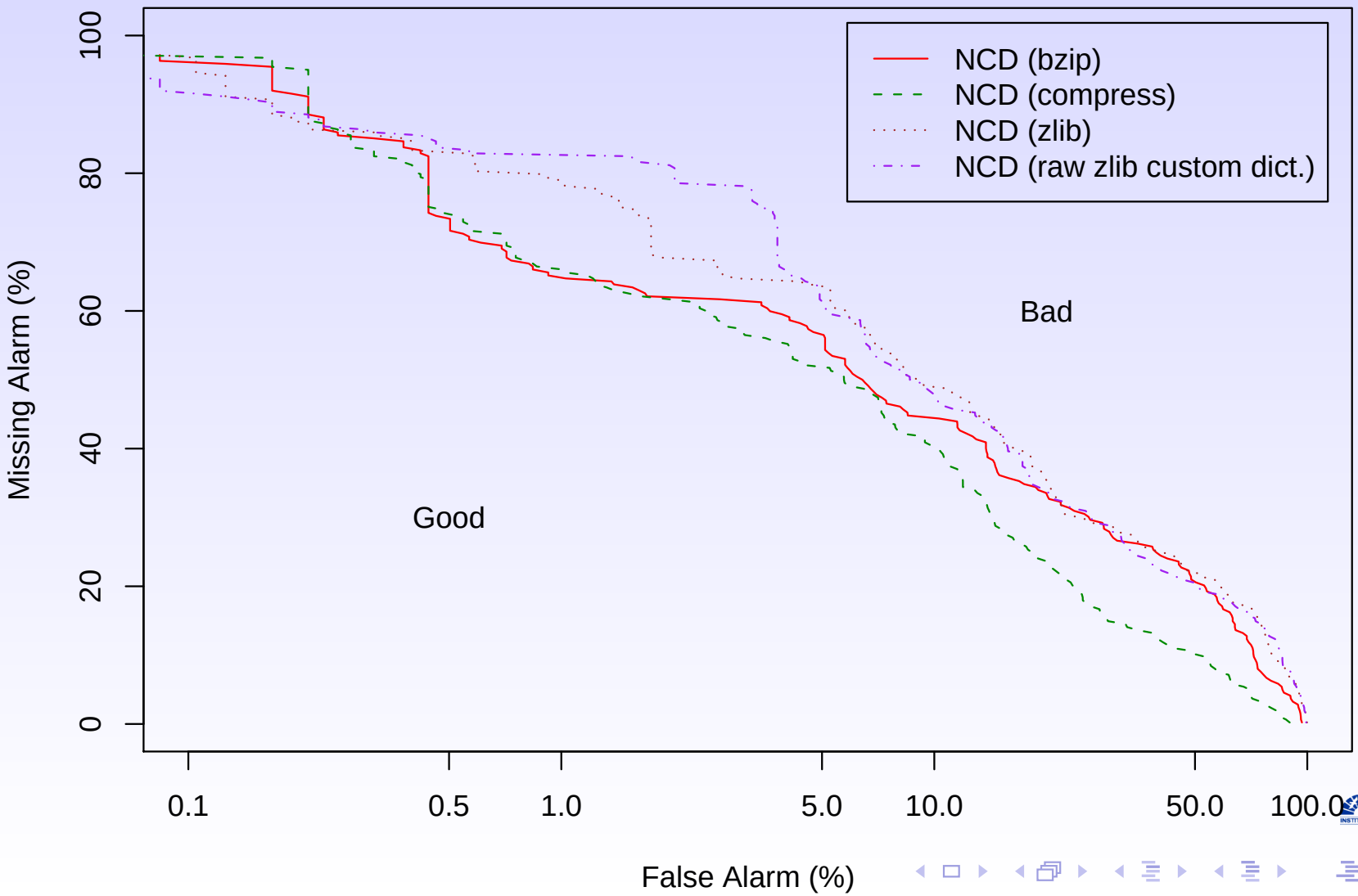
Método	*False Alarm Rate (%)*	*Missing Alarm Rate (%)*
Uniqueness	1.4	60.6
Bayes one-step Markov	6.7	30.7
Hybrid multistep Markov	3.2	50.7
Compression	5.0	65.8
Sequence-Match	3.7	63.2
IPAM	2.7	58.9
NCD (*bzip*)	**0.6**	**69.6**
NCD (*compress*)	**2.9**	**57.5**
NCD (*zlib*)	**1.2**	**77.9**
NCD (*raw zlib cust. dict.*)	**1.3**	**82.6**

Cuadro: Comparación de todos los métodos (apuntando a FAR de 1%)

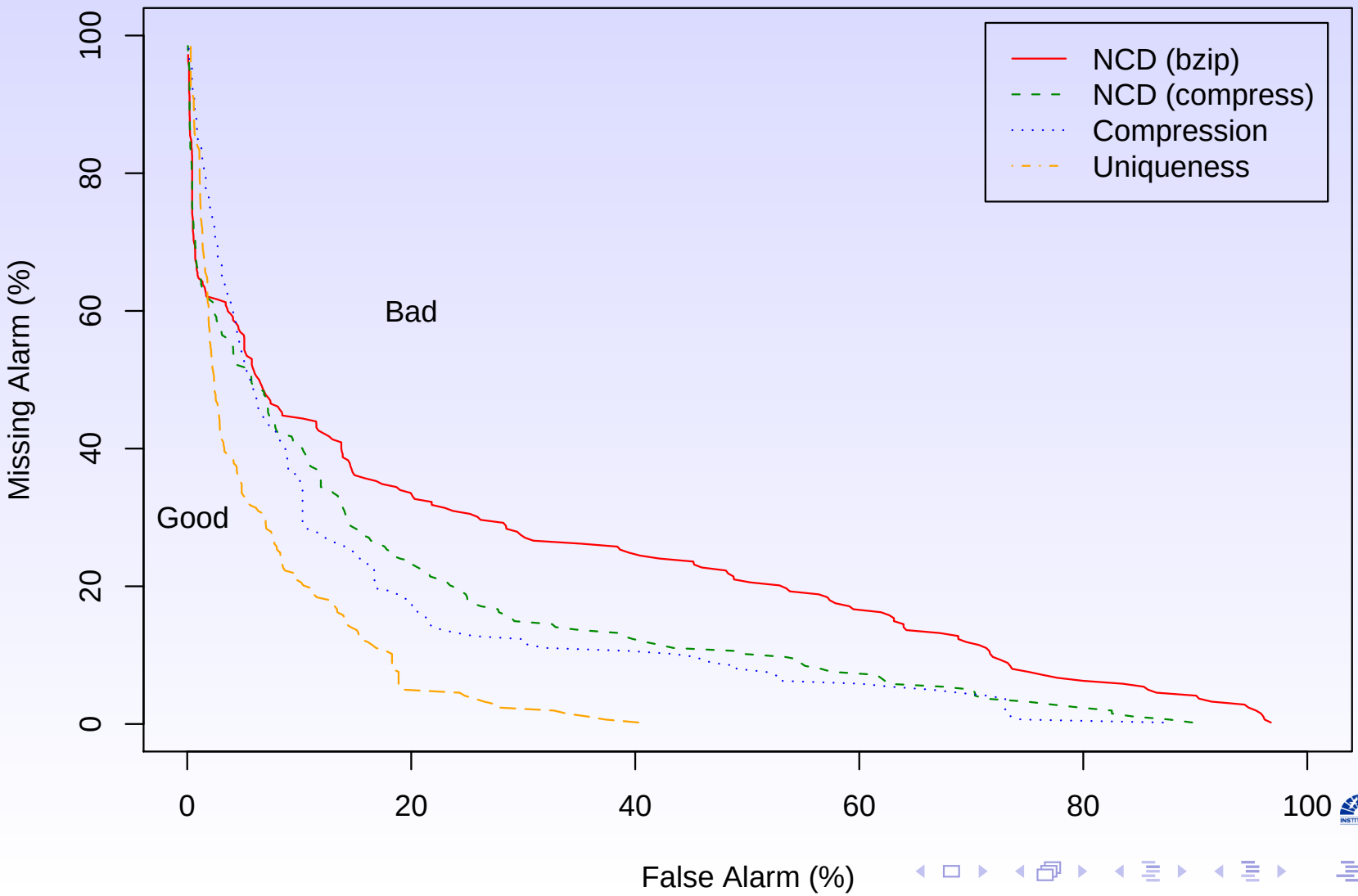
Curvas ROC (*NCD*)



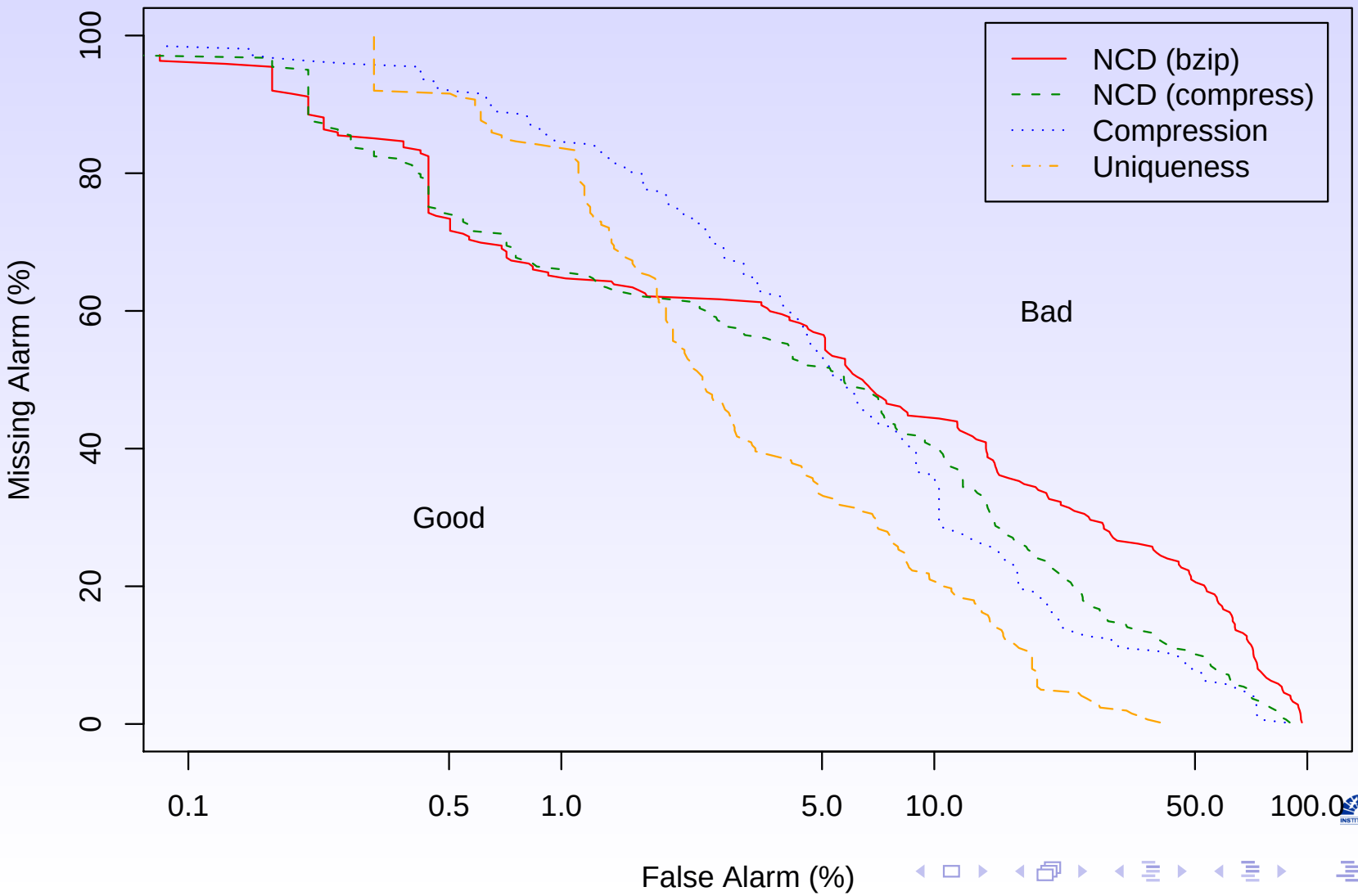
Curvas ROC (*NCD*)



Curvas ROC (Comparación)



Curvas ROC (Comparación)



Outline

Conclusiones y tareas pendientes

Conclusiones

- La técnica basada en *NCD*:
  - produjo resultados similares a los obtenidos por otras técnicas
  - produjo mejores resultados para *FAR* bajos
  - se acercó más a los valores de *FAR* deseados que otras técnicas
- El compresor `raw zlib` con diccionario no resultó en mejoras significativas

Conclusiones y tareas pendientes

Conclusiones

- La técnica basada en *NCD*:
  - produjo resultados similares a los obtenidos por otras técnicas
  - produjo mejores resultados para *FAR* bajos
  - se acercó más a los valores de *FAR* deseados que otras técnicas
- El compresor `raw zlib` con diccionario no resultó en mejoras significativas

Tareas pendientes

- Elección de un compresor óptimo para los datos
- Manejar bloques de datos de tamaño menor o variable
- Actualización de los "perfiles" de los usuarios
- Utilización de datos más completos

¿Preguntas?

Muchas Gracias

Maximiliano Bertacchini
mbertacchini@citefa.gov.ar

Distancias y métricas

Definición

D es una función de distancia en un conjunto Ω: $D : \Omega \times \Omega \rightarrow \Re^{+}$.

Definición

D es una métrica si satisface para todo $x, y, z \in \Omega$:

- $D(x, y) = 0$ iff $x = y$ (*identidad*)
- $D(x, y) = D(y, x)$ (*simetría*)
- $D(x, y) \leqslant D(x, z) + D(z, y)$ (*desigualdad triangular*)

Definición

D es una métrica admisible si es una métrica *computable* y *densa*.

Distancias y métricas

Definición

D es una función de distancia en un conjunto Ω: $D : \Omega \times \Omega \rightarrow \Re^+$.

Definición

D es una métrica si satisface para todo $x, y, z \in \Omega$:

- $D(x, y) = 0$ iff $x = y$ (*identidad*)
- $D(x, y) = D(y, x)$ (*simetría*)
- $D(x, y) \leqslant D(x, z) + D(z, y)$ (*desigualdad triangular*)

Definición

D es una métrica admisible si es una métrica *computable* y *densa*.

Ejemplo

$D(x, y) = |x - y|$ donde $x, y \in \Re^n$ (*distancia euclídea*)